

## Tema 1

## Generalidades

1.1 Definición.- La estadística es una ciencia que proporciona un conjunto de métodos que se emplean para recolectar, resumir, clasificar, analizar, e interpretar el comportamiento de los datos con respecto a una característica que se está estudiando o investigando. En primera instancia se aplica la estadística para describir el comportamiento de alguna característica en estudio y posteriormente se aplica para pronosticar el comportamiento de dicha característica.

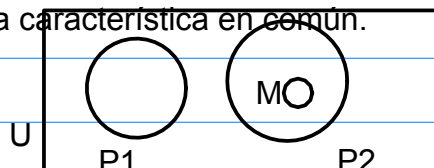
1.2. División de la estadística.- Se tiene dos aspectos bien marcados: el primero consiste en obtener datos, procesarlos y presentarlos (Estadística Descriptiva); y el segundo consiste en inferir o pronosticar (Estadística Inferencial).

1.2.1. Definición de estadística descriptiva.- es el conjunto de métodos que implican la recolección, presentación y caracterización de un conjunto de datos a fin de describir en forma apropiada las diversas características de estas. Es decir, que un estudio estadístico se considera descriptivo cuando solo se analiza y describe los datos.

1.2.2. Definición de estadística inferencial.- es el conjunto de métodos que posibilitan la generalización o toma de decisiones en base a una información parcial obtenida mediante técnicas descriptivas. Es decir, un estudio estadístico, se considera inferencial cuando pretende inferir o predecir conclusiones que atañen a toda la fuente de información de donde proviene los datos. Las predicciones se realizan con un grado o nivel de confianza que se conoce como probabilidad.

1.3. Población y Muestra.-

1.3.1. Población.- es el conjunto de todos los individuos, objetos u observaciones que poseen al menos una característica en común.



Ejemplos.-

- Las edades de los estudiantes de la EMI - UAC
- Las placas de circulación de los vehículos que forman parte del parque automotor de Cbba.

La población puede ser Finita o Infinita.

A) Población Finita.- es aquella que tiene un número limitado de elementos. P.ej. las estaturas de los estudiantes de la carrera de Ingeniería Civil de la EMI-UAC.

B) Población Infinita.- es aquella que no tiene límite, es decir, tienen un número infinito de elementos. P. ej. Peso de los adolescentes en Bolivia que se encuentran entre 12 y 18 años.

Parámetros.- Es una medida resumen que describe una característica de una población. P. ej. la estatura promedio de los estudiantes de la EMI-UAC.

Es posible obtener parámetros para poblaciones finitas. Pero, para poblaciones infinitas no es posible obtener parámetros precisos, por lo que se trata de inferir dichos parámetros.

1.3.2. Muestra.- Es una parte o un subconjunto de una población. Al proceso de obtener una muestra se denomina como muestreo.

ejemplo.-

Se realiza una encuesta pre electoral para las elecciones sub nacionales en el Cercado de Cbba. Con este fin se encuestan a 1500 personas registradas para votar y 400 personas votarán por el candidato A 500 por B 300 por C y 300 son indecisos.

Población: total de registrados en el Cercado de Cochabamba

Muestra: 1500 encuestados

Parámetros: votantes por cada candidato

#### 1.4. USOS DE LA ESTADÍSTICA.-

La estadística puede ser aplicada en todas las disciplinas, por ej. Salud y medicina, Economía, Sociales, Ingeniería.

La estadística se ha convertido en una herramienta muy importante de la investigación científica.

- a. Planteamiento del problema
- b. Recolección de datos
- c. Organización y clasificación de los datos
- d. Análisis e interpretación de los resultados
- e. Presentación de la información

##### a) Planteamiento del problema

Primero se debe identificar una situación problemática y se debe describir dicha situación.

- En qué consiste el problema objeto de investigación?

- Por qué se plantea su investigación?

- Dónde, cómo, quién, cuándo, se va a realizar la investigación

Proponer o formular el problema, éste tiene las variables que intervienen en una investigación que son: la variable dependiente y la variable independiente. P. ej.

El tratamiento actual de residuos sólidos en el Cercado de Cochabamba es inadecuado lo que genera focos de infección.

Una vez que se ha planteado el problema se propone una hipótesis, que es la solución al problema planteado.

##### b) Recolección de los datos.-

Una recolección correcta de datos es de gran importancia ya que la investigación estará basada en los datos recolectados. El hecho de que se obtengan datos inválidos esto generará resultados inválidos.

En esta etapa se debe preparar los instrumentos de recolección de datos, posteriormente se deben hacer pruebas piloto de los instrumentos, corregir, validar y ejecutar.

##### c) Organización y Clasificación de los datos.-

Se debe hacer un análisis de consistencia y ajuste de los datos. Se trata de asegurar la validez y confiabilidad de los datos recolectados. Luego se debe clasificar y tabular los datos para presentarlos en forma de cuadros, tablas y gráficos.

##### d) Análisis e interpretación de los resultados.-

La información de cuadros, tablas y gráficos se complementará con el cálculo de indicadores. A partir de todo lo anteriormente descrito el investigador debe llegar a unas conclusiones del trabajo que se ha realizado.

### 1.5. Recolección de datos

Se tienen diferentes formas para la recolección de los datos, entre ellas está:

A) El investigador puede recurrir a fuentes de información, estas fuentes pueden ser personas, instituciones. Estas fuentes pueden ser:

- Fuentes de datos internos, es decir que se encuentran al interior de la organización en la que se está desarrollando la investigación. Estas fuentes pueden ser: reportes financieros, reportes de operaciones (producción, ventas, logística investigaciones previas), Reportes especiales
- Fuentes de datos externos, son información generada por entidades externas a la organización que han realizado estudios estadísticos que son de interés para la organización.
- Fuentes primarias de información, cuando los datos son obtenidos directamente de la unidad de observación. p. ej. resultados obtenidos por censos, encuestas, cuestionarios, entrevistas.
- Fuentes secundarias, cuando se obtiene información estadística en base de los datos de las fuentes primarias.

B) Diseño de experimento, aplicación de modelos cuantitativos, cualitativos y explicativos, uso de laboratorios.

C) Encuestas y cuestionarios, dirigidas a una muestra representativa de la población. Pueden ser retrospectivas y prospectivas.

### 1.6. Tipos de Datos

Se puede obtener datos de carácter cuantitativo y cualitativo. Datos cuantitativos pueden expresarse por cantidades o números. Por otra parte los datos cualitativos pueden expresarse por atributos de cualidad.

a) Variable.- los datos obtenidos en una investigación están directamente relacionados con variables. Una variable es una característica de la población que se va a investigar y que puede tomar diferentes valores. Las variables se clasifican en:

- Cualitativas
- Cuantitativas

Las variables cualitativas están asociadas a una característica cualitativa, cuyos valores son cualidades. P. ej. Profesión, religión. Las variables Cualitativas pueden ser ordinales o nominales.

Las variables cualitativas ordinales denotan jerarquía. P. ej. Grado de instrucción.

Las variables cualitativas nominales no denotan jerarquía, es decir no importa el orden. P. ej. sexo, religión, profesión.

Las variables cuantitativas están asociadas a una característica que se puede medir y expresar esta dimensión por números o cantidades. Las variables cuantitativas pueden ser continuas o discretas. Son variables cuantitativas continuas cuando el valor de la variable puede ser expresado en fracciones o decimales. P. ej. peso 70,3 kg. estatura: 1,74 m.

Por otra parte, las variables cuantitativas discretas, son expresadas por valores enteros. P. ej. Número de hijos por familia. Edad en años cumplidos: 45, 20

## 2.1. PRESENTACION DE DATOS

Un conjunto de datos debe ser presentado de forma adecuada, dependiendo del tipo de variables y de la cantidad de los mismos.

Hay tres formas de presentar datos:

- A) Ordenación de datos
- B) Tabulación discreta
- C) Clasificación de datos

2.1.1. Ordenación de datos, o fila de datos, es un conjunto de datos sin clasificar y pueden estar presentados de mayor a menor o de menor a mayor.

P. ej. Ordenación de datos de las edades de los estudiantes de CITEC 2A, EMI CBA.

18, 18, 18, 18, 18, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 20, 20, 20

2.2.2. Tabulación Discreta, es una tabla en la que se tiene dos columnas. En la primera columna se tiene los diferentes valores que adopta la variable en estudio, y en la segunda columna se tiene la frecuencia que representa al número de veces que se repite un dato. Se llama Tabulación discreta porque generalmente se trabaja con variables discretas.

P. ej. Se presenta las edades en años cumplidos de los estudiantes de la carrera de ingeniería civil, de la EMI CBA.

$X_i$  = las edades de los estudiantes de la carrera de ing. civil de la EMI CBA [años cumplidos]

$X_i$	$f_i$	
19	20	$n$ = frecuencia total
20	25	
21	18	$f_i$ = frecuencia simple absoluta
22	30	
23	35	
24	30	
25	32	
26	10	
$n =$	200	

2.2.3. Clasificación de datos, Es un resumen de todos los datos presentados a través de intervalos que son denominados como clases o categorías. También debe tener una columna de frecuencia absoluta simple. Generalmente se emplea esta forma de presentación para variables cuantitativas continuas, pero también se puede emplear esta forma con variables discretas. Esta forma de presentación debe ser utilizada cuando trabajamos con cantidades grande de datos.

P. ej. Se presenta las estaturas en cm de los estudiantes de la EMI CBA

	Intervalos de clase [Linf, Lsup]	f Frecuencia de clase
clase 1	[156, 160[	200
clase 2	[160, 164[	300
clase 3	[164, 168[	1300
clase 4	[168, 172[	1200
clase 5	[172, 176[	700
clase 6	[176, 180[	300
		n = 4000

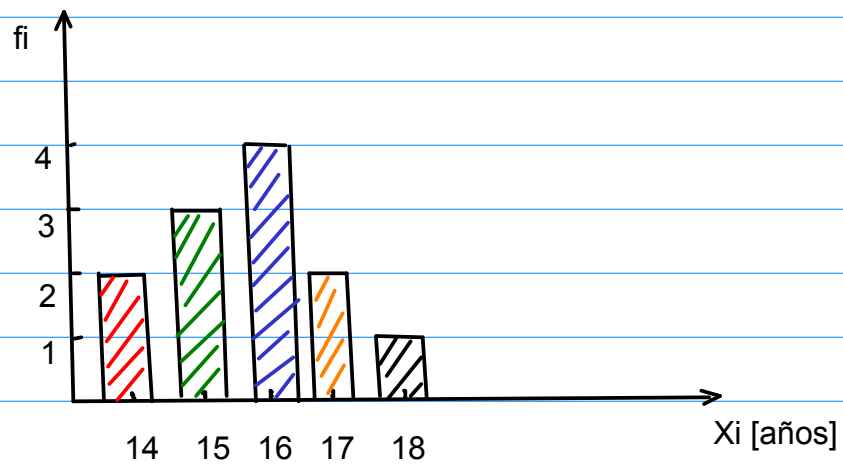
k = el número de clases diferentes que tiene una clasificación de datos.

## 2.2. DIAGRAMAS

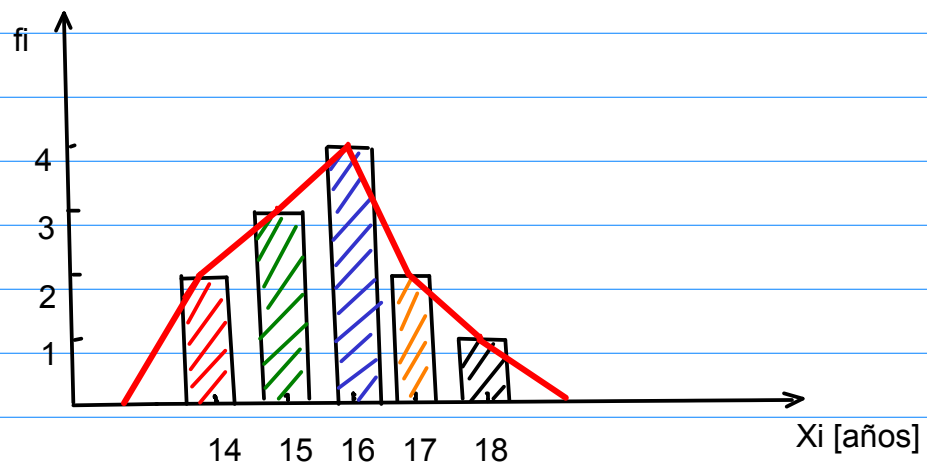
### A) DIAGRAMAS DE BARRAS

para tabulación discreta:

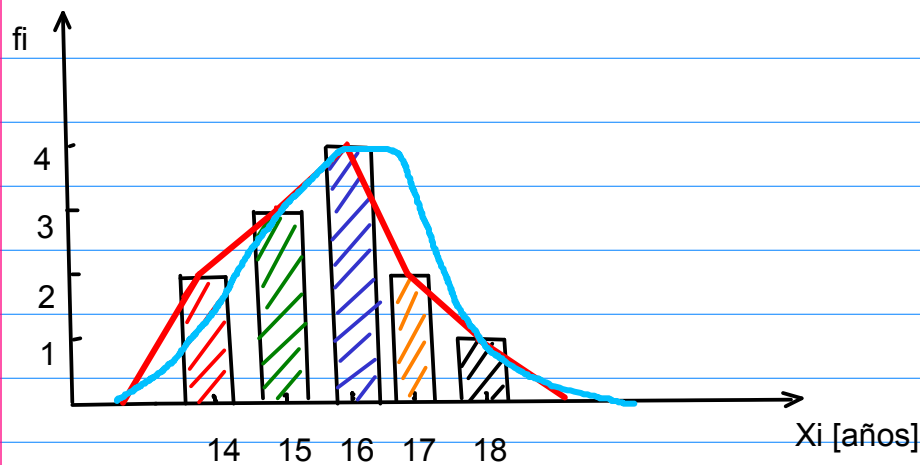
$X_i$	$f_i$
14	2
15	3
16	4
17	2
18	1



### B) POLÍGONO DE FRECUENCIAS



### C) CURVA DE FRECUENCIAS

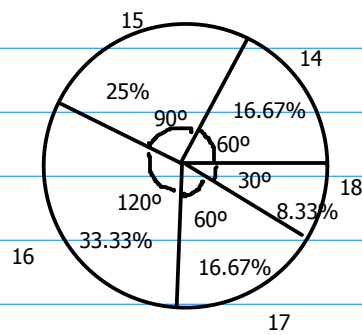


### D) DIAGRAMA DE SECTORES (TORTAS)

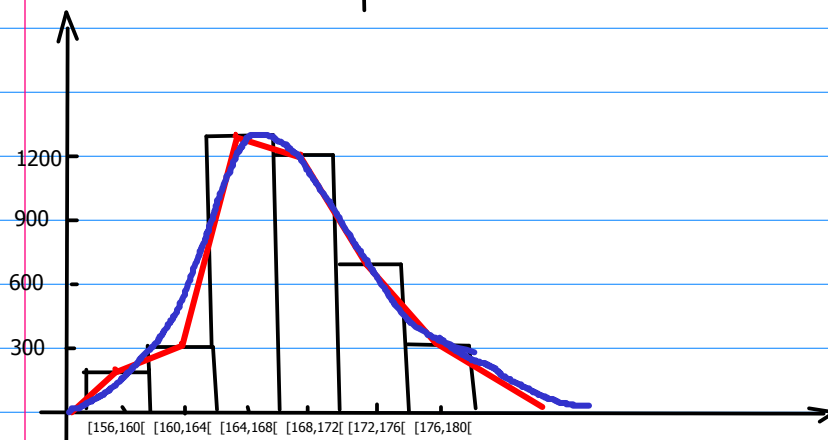
$$\alpha = \frac{f_i}{n} \times 360^\circ$$

Xi	fi	$\alpha$
14	2	60°
15	3	90°
16	4	120°
17	2	60°
18	1	30°

n = 12



	Intervalos de clase [Lin. Lsup]	f Frecuencia de clase
clase 1	[156, 160[	200
clase 2	[160, 164[	300
clase 3	[164, 168[	1300
clase 4	[168, 172[	1200
clase 5	[172, 176[	700
clase 6	[176, 180[	300
		n = 4000



### 2.3. FRECUENCIAS RELATIVAS

Las frecuencias relativas permiten la comparación de datos usando proporciones y porcentajes.

Se llama Frecuencia Relativa h, al valor de la división de una Frecuencia Absoluta Simple f entre la Frecuencia total n.

Cuando la frecuencia relativa se multiplica por 100, se obtiene la Frecuencia Porcentual p.

P. ej.

X	f
14	2
15	3
16	4
17	2
18	1
<u>n = 12</u>	

$$h = f / n$$

$$p = h * 100$$

X	Frecuencia absoluta f	Frecuencia relativa h	Frecuencia porcentual p
14	2	$2/12 = 0.17$	<u>17 %</u>
15	3	$3/12 = 0.25$	25 %
16	4	$4/12 = 0.33$	33 %
17	2	$2/12 = 0.17$	17 %
18	1	$1/12 = 0.08$	8 %
	<u>n = 12</u>	<u>1</u>	<u>100 %</u>

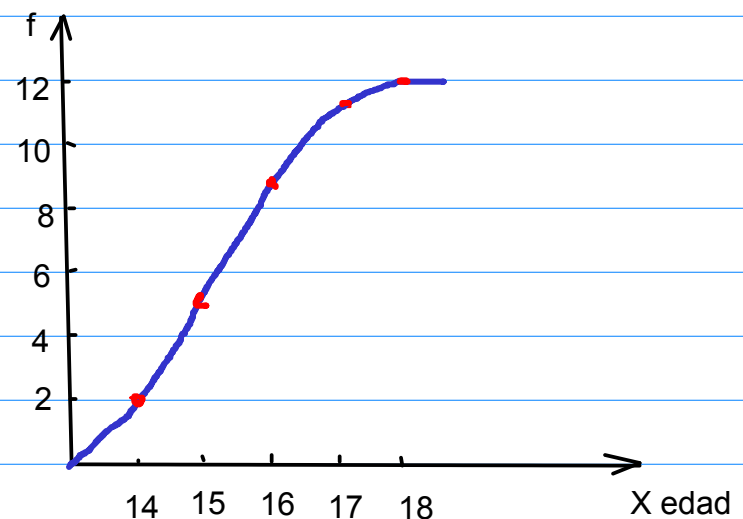
### 2.4. FRECUENCIAS ACUMULADAS

#### A) FRECUENCIA ACUMULADA MENOR QUE (F)

Se llama Frecuencia Acumulada Menor Que a la suma de todas las frecuencias correspondientes a los datos menores o iguales al que se considera, en un conjunto de datos agrupados.

P. ej.

X	f	F
14	2	2
15	3	2+3=5
16	4	2+3+4=9
17	2	2+3+4+2=11
18	1	2+3+4+2+1=12

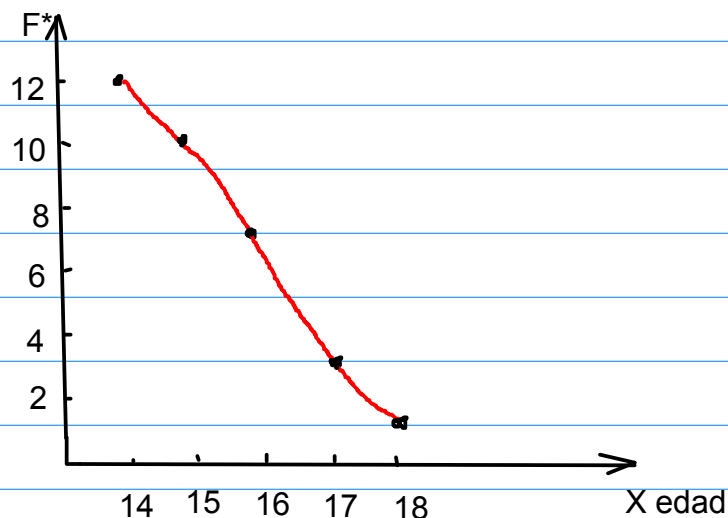


## B) FRECUENCIA ACUMULADA MAYOR QUE ( $F^*$ )

Se llama Frecuencia Acumulada Mayor Que a la suma de todas las frecuencias correspondientes a los datos mayores o iguales al que se considera, en un conjunto de datos agrupados.

P. ej.

X	f	$F^*$
14	2	$2+3+4+2+1=12$
15	3	$3+4+2+1=10$
16	4	$4+2+1=7$
17	2	$2+1=3$
18	1	1



Tanto para la Frecuencia acumulada menor que y la Frecuencia acumulada mayor que se generalizan los conceptos de Frecuencia Relativa.

	X Datos	Frecuencias Simples			FRECUENCIAS ACUMULADAS					
					Frec. Ac. Menor Que			Frec. Ac. Mayor Que		
		f	h	p	F	H	P	$F^*$	$H^*$	$P^*$
1	14	2	0.17	17%	2	0.17	17 %	12	1.00	100 %
2	15	3	0.25	25%	5	0.42	42 %	10	0.83	83 %
3	16	4	0.33	33%	9	0.75	75 %	7	0.58	58 %
4	17	2	0.17	17%	11	0.92	92 %	3	0.25	25 %
5	18	1	0.08	8%	12	1.00	100 %	1	0.08	8 %
		n=12	1	100%						

En una tabulación distcreta K representa al número de datos diferentes que adopta X. Ej :  $K=5$  ←

### EJERCICIO

X	f
1	12
2	16
3	20
4	14
5	10
6	8



## CLASIFICACIÓN DE DATOS

Se entiende por clasificar al procedimiento de distribuir los datos en clases, estableciendo intervalos, para de esta manera facilitar el análisis de los datos.

X				
	Clases	Amplitud de clase	Marca de clase	f
1	[10,20[	10	15	21
2	[20,30[	10	25	9
3	[30,40[	10	35	0
4	[40,50[	10	45	14
				<u>44</u>
n = 44				

Intervalos o clases, [l inf, l sup[

Amplitud o ancho de clase  $c = l \text{ sup} - l \text{ inf}$

Marca de clase =  $(l \text{ sup} + l \text{ inf}) / 2$ , es el valor intermedio del intervalo y que generalmente representa a la clase.

$$n = \sum f_i = \text{Frecuencia total}$$

K = número de clases

$$\text{Rango} = R = X \text{ max} - X \text{ min} = 50 - 10 = 40$$

$$K = \sqrt[n]{n} = \sqrt[40]{44} = 6.63 = 7$$

$$R = c * K \quad c = R/K$$

Ej: Los siguientes valores representan los salarios expresados en dólares de un grupo de trabajadores:

102, 110, 120, 130, 135, 140, 145, 145, 150, 155, 160, 170, 170, 178

$$n = 14$$

$$K = \sqrt[14]{14} = 3.74 = 4$$

$$X \text{ max} = 178$$

$$X \text{ min} = 102$$

$$R = 178 - 102 = 76$$

$$c = 76/4 = 19 \approx 20 \quad R' = 20 * 4 = 80 \quad R' - R = 80 - 76 = 4/2 = 2$$

clases		$X \text{ min} - 2 = 102 - 2 = 100$	$X \text{ max} + 2 = 178 + 2 = 180$
[102,121[	3		
[121,140[	2		
[140,159[	5		
[159,178[	3		
<u>13</u>			

102. 110. 120. 130. 135. 140. 145. 145. 150. 155. 160. 170. 170. 178

| clases    | ancho de clase | MÁrca de clase | FRECUENCIAS SIMPLES |      |    | FREC. AC. MENOR QUE |      |     | FREC. AC. MAYOR QUE |      |     |
|-----------|----------------|----------------|---------------------|------|----|---------------------|------|-----|---------------------|------|-----|
|           |                |                | f                   | h    | p  | F                   | H    | P   | F*                  | H*   | p*  |
| [100,120[ | 20             | 110            | 2                   | 0.14 | 14 | 2                   | 0.14 | 14  | 14                  | 1    | 100 |
| [120,140[ | 20             | 130            | 3                   | 0.21 | 21 | 5                   | 0.35 | 35  | 12                  | 0.86 | 86  |
| [140,160[ | 20             | 150            | 5                   | 0.36 | 36 | 10                  | 0.71 | 71  | 9                   | 0.64 | 64  |
| [160,180[ | 20             | 170            | 4                   | 0.29 | 29 | 14                  | 1    | 100 | 4                   | 0.29 | 29  |

14      1